

Robotic Sensor-Motor Transformations

**Michael R. Blackburn and
Hoa G. Nguyen**

Naval Command, Control and
Ocean Surveillance Center
Research, Development, Test, and
Evaluation Division
San Diego, CA 92152-7383

Abstract

This paper summarizes work performed at NRaD during FY94 on the integration of robotic sensor and motor systems¹. Two robotic applications were involved. These are the visual control of a five degrees of freedom manipulator arm in three dimensions, and the visual control of a mobile platform for target acquisition, tracking and trailing with obstacle avoidance. Solutions to both applications were constrained by the intentional restriction that information could only be gained autonomously by the robot through vision and proprioception (internal information on mechanical position). Furthermore, the only visual information allowed was the frame-to-frame image flow derived from the contrast gradients. Both reflexive (reactive) and adaptive (learning) algorithms were studied.

1 Introduction

The overall objectives of this research are to design, develop, and test an artificial autonomous visually guided motor system by using adaptive neural network computer algorithms that explicitly emulate the functional architecture of known and hypothesized biological mechanisms. The goal is to improve target recognition and discrimination under different transformations of the target image. The methods that we are exploring to improve automatic target recognition involve active perception and self-determined manipulation of the target, or of the robot platform relative to the environment.

¹This research is supported by the Advanced Research Projects Agency and the Office of Naval Research under contract number N0001493WX2D002.

This project expands on earlier Navy work in autonomous active machine vision [Blackburn et al., 1987; Blackburn and Nguyen, 1990]. The machine vision system is being adapted to a stereo pan, tilt and vergence mechanism, and used to control a five degree-of-freedom stationary robotic manipulator arm, and a mobile robot platform. The project combines vision, navigation and manipulation to study learning of cross-modal sensor-sensor and sensor-motor sequences. Mechanisms to demonstrate a motor (behavioral) criterion of scale and rotation invariant recognition are being developed. In addition, the visual-motor algorithms are being applied on the mobile robot to the problem of target discrimination, tracking, and trailing while on the move.

The algorithms of active perception employ both reflexive and adaptive mechanisms. Reflexive mechanisms provide low level, generic, and fault tolerant solutions to problems such as target detection, segmentation, and obstacle avoidance, while adaptive mechanisms provide intrinsically modifiable solutions to difficult problems such as eye-hand calibration and target discrimination.

2 Visual Control of a Mobile Robot

Figure 1 shows our mobile robot under visual control trailing a walking human target. The circumstances that challenge this task include the complexity of the background, the proximity of the target and the velocity of its image on the visual field, the absence of human assistance or intervention, the absence of unique distinguishing features associated with the target, and the limitations of on-board processing power and energy resources. Details of the algorithms and hardware used in this study are reported in a technical paper contained in this volume [Blackburn and Nguyen, 1994a].

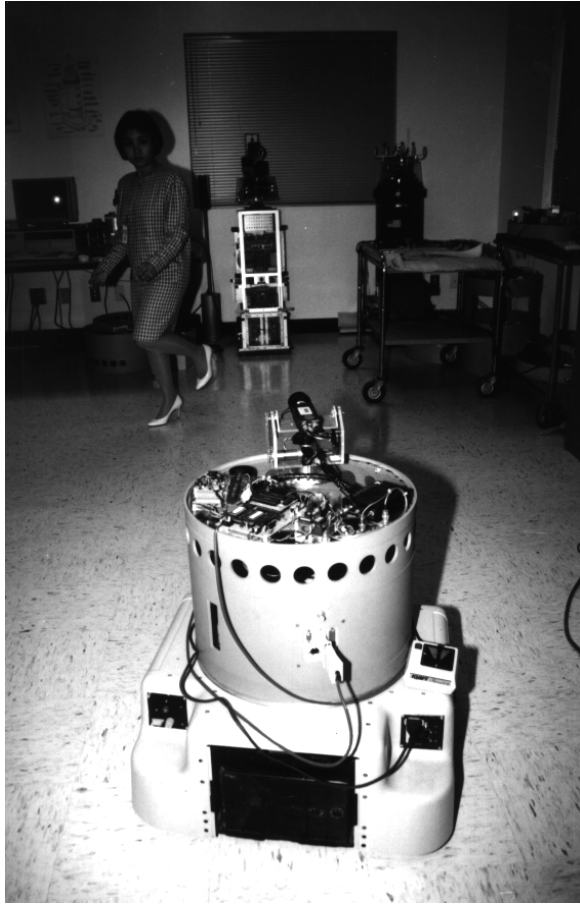


Figure 1. Autonomous visually controlled mobile robot trailing a walking human in a visually cluttered environment.

The idea is to solve a difficult problem in a parsimonious way following the examples of nature. We assume that natural selection favored efficiency. The need for parsimony in robotics is also related to economics. For robots to be considered viable and accepted into the workplace they must be cost effective as well as competent.

At the present the robot vision system is the only means by which we allow the robot to gain the necessary information about its external environment. To further restrict the nature of the available information, we only analyze the image flow from the contrast gradients. Thus, motion information alone is available. The autonomous robot must use this information for multiple purposes. First the robot must acquire and maintain a target. A potential target is detected by motion in the peripheral visual field. Target acquisition is accomplished through saccades and smooth pursuit motions utilizing a pan and tilt

mechanism. The target then is defined as any source of motion that has been placed on the central region of the receptor surface. Second the robot must assess the behavior of the target and respond appropriately. This is the tracking and trailing task. If the motion in the central region of the receptor surface is contracting, the target is assumed to be receding and an approach response is triggered. If the motion is expanding, the target is probably looming and the robot's forward motion is suspended. Lastly, the robot must recognize and respond appropriately to non-target objects (i.e. obstacles) without losing sight of its target. Obstacles are defined by any motion on the peripheral receptor field after a target has been acquired. Other researchers have divided these objectives among different sensor modalities, thereby simplifying the task that must be accomplished by each sensor system. We accept our limitations, however, to explore the full potential of vision in the control of the mobile robot.

Two significant related problems were encountered in this work. First is the maintenance of the target on the center of the visual field. Second is the detection of the moving target while the robot itself is moving through a visually complex environment. Only partial solutions to both problems have been achieved.

We use a receptor surface that has an inhomogeneous resolution. Like the biological retina, we incorporate a high resolution fovea and a peripheral retina whose resolution decreases as a function of eccentricity. A log-polar transformation is used to sample, integrate, and map the receptor input to our computational plane. This architecture is an efficient means of data compression as only the central region is analyzed in detail. There are other advantages as well. The larger receptive fields in the periphery integrate contrast changes, increasing sensitivity to jiggle, and provide a larger separation for sampling the higher velocities that are expected there.

In a mechanism such as ours that depends on target motion for detection, acquisition and evaluation, the successful localization and pursuit of the target means that the target is stabilized on the center of the receptor surface. Stabilization occurs when the target motion is minimized, and this eliminates the information upon which the target is maintained so that the target may disappear. Due to the inhomogeneous resolution of

the receptor surface, the central region has a greater sensitivity to slow motion than does the peripheral region. Targets located at a distance may be pursued by the vision system without generating competing motion in the periphery. However, when the vehicle is underway, the self-induced optic flow is problematical to smooth pursuit of a slowly moving target.

The second problem results from the competitive mechanisms that select the best candidate for a target. Motion again is the criterion. A moving platform creates relative motion on its sensor surface in the presence of visual contrast from both foreground and background objects. This induced motion tends to be correlated however and thus may be predictable with self awareness of self motion. A moving target on the other hand can have an unpredictable velocity on the sensor surface. We developed a motion segmentation algorithm, based on a biological model, that takes advantage both of predictability and of local consistency. Local center surround mechanisms on the motion field reduce the effects of correlated motion, and enhance unique motion. Furthermore, self-awareness of auto-motion in one direction can be used to reduce sensitivity to optic flow in the opposite direction. This inhibitory process conflicts, however, with the mechanism of target maintenance on the central region because it eliminates necessary feedback for pursuit velocity. Therefore, the central region has been exempted from auto-motion inhibition.

While on the move, the robot vision system can detect targets that are also in motion. The mechanism of motion segmentation favors the unique motion of the target. Additionally, the network is almost never without a target. The camera is attracted to static objects during auto-motion. These are only ignored after stabilization when they fail to further move. Camera stabilization during auto-motion further reduces the induced optic flow of the static background, favoring the acquisition of animate targets.

The stabilization mechanism is as yet unreliable in our implementation because of the poor performance of the smooth pursuit system. Currently, the maintenance of a moving target from a moving platform is more successfully performed by small corrective saccades because the pursuit system cannot keep up with the induced velocities.

For obstacle avoidance an advantage is gained with the velocities that accompany auto motion. Nearby obstacles result in easily detected flow on the peripheral retina. The lateral imbalance in this flow is used to maneuver the robot around the obstacles.

The algorithms for visual motion analysis and robot control were sufficiently efficient to run on the robot at eight frames per second using a pair of i860 co-processors in parallel with an 80486 PC located onboard the mobile robot. Much of this time is consumed in importing the image frame to the i860s from the frame grabber and in the definition of the local receptive fields. The use of a vision chip that could perform the data reduction of the log-polar mapping prior to digitization would greatly increase frame rates.

To improve trailing performance we need better methods to maintain the fix or attention on a moving target from a moving robot while avoiding obstacles. Some memory for the target's position and behavior may allow recovery of the target after the vehicle has been deflected by an obstacle. Some pattern processing that would uniquely mark the target could also help to maintain attention.

The ability to detect and track a moving target while the robot platform is itself on the move has many applications among which are autonomous automobile navigation, battlefield unmanned forward observers, and automated factory or warehouse material distribution. The common element in all of these applications is the relief of a human operator from the burden of maintaining attention to the tasks of target acquisition and maintenance, and of navigation through a complex and unpredictable environment.

3 Robot Visual Control of a Manipulator Arm in Three Dimensions

Figure 2 shows our mobile robot parked before a manipulator arm, over which it has assumed visual control. The problem addressed in this configuration is the visual direction of the manipulator arm following a process that learned the inverse kinematics model. This is a difficult problem because the solutions can be non-unique, the information available can be noisy, and because the system calibration can change over time. Several researchers have already addressed similar problems, developing adaptive algorithms to learn the inverse kinematics [Kuperstein and Rubinstein,

1989; Martinetz and Schulten, 1990; Li and Ogmen, 1994]. We have made modest improvements to these methods that enhance learning rates and accuracy, while reducing computational complexity. The details of our work on this problem are also available in a technical report contained in this volume [Blackburn and Nguyen, 1994b].

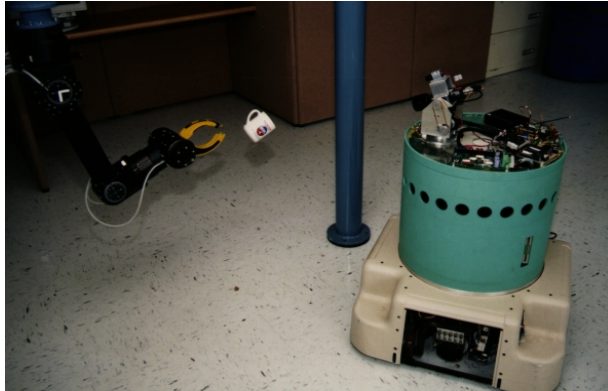


Figure 2. Adaptive autonomous visual control of a robot manipulator arm.

Work in three dimensions requires information on target position in X, Y and Z. A single camera can provide in a straight forward way the information on X and Y, and through active perception, information on the relative depth of objects, based primarily on motion parallax and occlusions. Two cameras in binocular vision provide additional information on absolute depth through vergence measures and on relative depth through location disparities in the projections. Primate vision systems acquire and locate targets and direct arm motion using both binocular and monocular vision, plus motion analysis. The architecture of the retina and the log-polar mapping of the receptor surface to the visual cortex permit a simple analysis of depth or relative positions of objects in the visual field. The vergence that results from the fixation of both eyes on a selected target eliminates the binocular disparity of the target, but allows the assessment of the location of objects in the vicinity of the target. One such object of great importance for our purposes is the end-effector of the manipulator arm as it reaches for the target. We use pan/tilt/vergence information of a foveated target to locate the target in the 3D coordinate system of the robot and direct the ballistic phase of the robot manipulator arm end-effector to the target location. Then in the final phase of reaching, we use the relative motion of the end-effector on the two retinæ to correct any reaching errors and bring the

end-effector exactly onto the target. The control parameters of both phases of reaching to a target are learned by the robot through experience.

Four adaptive neural network algorithms were developed and compared for stereo control of reaching using a simulated 3 degree of freedom manipulator arm. The network architectures included the standard three layer perceptron with Back Propagation learning, a two layer perceptron with preprocessing using vertex-normal features, a Kohonen self-organizing map, and an associative mapping of distributed representations of manipulator and camera joint space with population coding. The performances of all adaptive algorithms were superior to a look-up table given the same numbers of exemplars. Reaching accuracy differed among the algorithms, but was primarily a function of network complexity and training time. The most efficient algorithm was the two layer perceptron with vertex-normal feature preprocessing. The vertex-normal feature preprocessing eliminated the need for one adaptive layer (the hidden layer) that is requisite for non-linear mappings in the three layer perceptron. Consequently, back propagation of error learning was not required, and the permitted use of the simpler delta rule learning greatly increased learning rates and run-time adaptability.

The addition of a second error correction strategy that involved learning of velocity correlations under continuous visual feedback reduced errors to an arbitrarily small degree and obviated the need for either large networks, large numbers of exemplars, or large training times.

The mobile robot manipulator arm system of Figure 2 presents difficulties in initial calibration as well as in the maintenance of calibration. It is not efficient to learn the calibration anew each time the robot rolls up to the arm. We need to develop adaptive calibration methods for arbitrary configurations of the vision system and manipulator arm. To accomplish this the control algorithms could search the visual space for frames of reference and compute the transformations of the image data needed for invariant association with the motor output.

While waiting for a stereo vision pan, tilt and vergence mechanism to arrive from the manufacturer, we developed an active laser/vision triangulation mechanism for target depth discrimination. This system requires saccades to

each point in space from which absolute depth information is desired. The returned data substitute for the missing vergence information of a binocular system, but does not directly provide the relative depth information necessary to visually servo the arm end-effector into the target.

4 Invariant Recognition through Active Perception

Sensor-motor integration, in one form commonly known as eye-hand coordination, is a process that permits the system to make and test hypotheses about objects in the environment. In a sense, nature invented the scientific method for the nervous system to use as a means to predict and prepare for significant events. A reactive organism must depend on speed to survive, but a predictive system can avoid problems altogether.

The motor component of perception compensates for an uncooperative environment. Not only does the use of effectors provide mobility, but it alters the information available, uncovering new opportunities to exploit. Random motion can achieve this, but at a cost in energy expenditure and at the cost of opening the host to exploitation. Neither does random motion permit the testing of the spatial relationships of information. The development of purposive movement allows the host to judiciously act in the environment and sample the results. Prediction forms the basis of the judgement to act, and the results are used to formulate new predictions. The successful match of prediction and results, just as in the scientific method, increases certainty of the validity of the hypothesis or the prediction and can lead to new predictions and new associated actions. An action-sensation-prediction-action chain is established through experience and conditioned learning that allows the organism to efficiently meet its metabolic needs, survive and procreate.

We demonstrated previously how an artificial vision system could learn such a sequence [Blackburn and Nguyen, 1990]. One behavioral piece of evidence for the action-sensation-prediction sequence is the scan path. The scan path is a sequence of eye (or camera) saccades that sample a target in a regular way to collect information. Figure 3 shows scan paths that were produced by our artificial vision system before and after a period of learning. The path in (a), made before learning, is the result of random saccades to regions of high information (defined by image

complexity). The path in (b) resulted from the interaction of learned expectation and the available features. The scan path after learning became more regular and the inter-saccade interval was reduced compared to the naive state. The sequential visits to the three locations on the sailboat represent the "recognition" of the image and its associated appropriate behavior in the most simple sense.

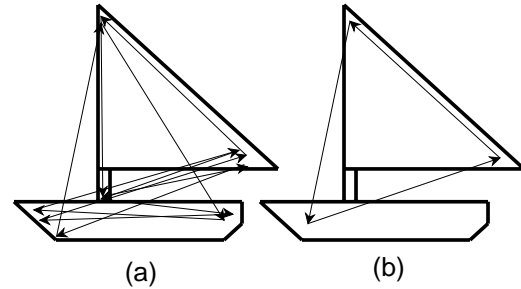


Figure 3. Scan paths of a naive network (a) and of an experienced network (b) to a static line drawing of boat. Arrows indicate direction of saccades. New target regions of the image were detected by small oscillations of the receptor surface, and selected competitively in the superior colliculus model network. Learning, based on experience scanning the image, provided a bias from the cortex to the superior colliculus that favored the regions of visual space from which the most likely (expected) saccade targets could be selected.

We propose that to achieve invariant recognition, an appropriate behavior must be transformed by the image parameters. For the example of a scan path behavior to an image that is presented at different sizes, the saccade amplitudes must be modulated. This could be accomplished by the use of a scale factor derived from some independent measure of image size, however, the topographical mapping that is common in the nervous system permits a natural rescaling of saccade amplitude based upon the locus of activity on the output map. To change the locus of activity, it is only necessary to match the expectation from the associative map with the available sensor information. It is the expectation that must be size invariant. We approached this by developing a progression of abstract processing stages that integrated features from lower stages to the most abstract relationships that are then stored in associative memory [Blackburn, 1992]. This pathway is paralleled by a progression of processing stages in reverse that accomplishes feature reconstruction. The information is passed outward, providing an increasing degree of spatial specificity when gated by the forward sequence of feature integrators.

Figure 4 shows the development of a scan path to a simple square that is changing size. For the individual network that experienced the moving square, the activation of complex features in the highest associative processing layers resulting from the selected locations on the image would constitute its invariant perception of a square. The evidence that squares of different sizes are perceived similarly is that the scan paths are similar.

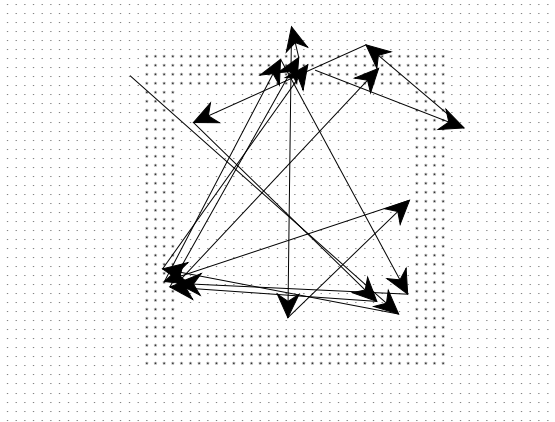


Figure 4. Simulated scan path of a square during changes in size. In this example the size of the square oscillated three times over a range of four hundred percent. Each arrowhead represents the visual rest point between saccades. The vision system developed this scan path after a period of learning in which motion and pattern information were combined to predict the changes in features that occurred as the image changed in size. The concentration of rest points, due to repeated visits to regions of the square, indicate a non-random scan of the image or one that resulted from a learned expectation of what the square looked like irrespective of size.

The key mechanism in our model of this process is the integration of pattern and motion information. As information moves deeper through the processing stages the size parameters become less specific (or more invariant) while the motion parameters become more specific. The importance of specific motion information at high levels in the visual processing is to predict changes in observable features given the behavior of the target. For example, by experience we have learned that two eyes should appear when a face turns to meet us. The observed motion of the head cues the expectation of new features. When these features appear, the face is recognized.

Even a target that has not moved during the scan can be made to reveal critical information by

deliberate motion of the observer in relation to it. Humans deal with most inanimate targets in this way. Objects are viewed from different perspectives, or picked up and examined. We plan to use the manipulator arm in just this role. Under visual control, the arm will be directed to grasp, rotate, or otherwise justify an object to the perceptual expectations of the machine vision system. This work is in progress.

References

Blackburn, M.R. [1992] Response invariance in an artificial visual-motor system. In IR-IED '92 Annual Report, NRaD TD 2412, October 1992, NCCOSC-RTD&E Div., San Diego, CA 92152-5001, 9-26.

Blackburn, M.R. and Nguyen, H.G. [1990] Modeling the biological mechanisms of vision: Scan paths. In Avula, X. (ed.) *Mathematical and Computer Modeling in Science and Technology*. Oxford: Pergamon Press, 311-316.

Blackburn, M.R., and Nguyen, H.G. [1994a] Autonomous Visual Control of a Mobile Robot. In Proceedings of the Image Understanding Workshop. Monterey, CA November 1994.

Blackburn, M.R. and Nguyen, H.G. [1994b] Learning in robot vision directed reaching: A comparison of methods. In Proceedings of the Image Understanding Workshop. Monterey, CA, November 1994.

Blackburn, M.R., Nguyen, H.G. and Kaomea, P.K. [1987] Machine visual motion detection modeled on vertebrate retina. SPIE Proceedings 980, 90-98.

Kuperstein, M. and Rubinstein, J. [1989] Implementation of an adaptive neural controller for sensory-motor coordination. IJCNN 89 Proceedings, II, 305-310.

Li, L. and Ogmen, H. [1994] Visually guided motor control: Adaptive sensorimotor mapping with on-line visual-error correction. Proceedings of the World Congress on Neural Networks, June 5-9, 1994, San Diego, CA, II, 1127-134.

Martinetz, T.M. and Schulten, K.J. [1990] Hierarchical neural net for learning control of a robot's arm and gripper. IJCNN 90 Proceedings, II, 747-752.